Downlink Resource Allocation for On-Demand Multimedia Services in Cellular/Infostation Integrated HSR Networks

Yan Lei, Gang Zhu, Chao Shen, Ning Zhang, Jingjing Xu, Lijun Jiang

The State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

11111007@bjtu.edu.cn; gzhu@bjtu.edu.cn; shenchao@bjtu.edu.cn; 13120177@bjtu.edu.cn; 09111052@bjtu.edu.cn; ljjiang@bjtu.edu.cn

Abstract

In this paper, we investigate the downlink resource allocation problem for on-demand CM services in HSR OFDMA systems with a cellular/infostation integrated network architecture. The resource allocation problem is formulated by a two-stage optimization programming, which aims at maximizing the total reward of delivered services then minimizing the weight total number of cumulative discontinuity packets over the trip of the train. An equivalent one-stage programming is proposed to resolve the difficulty of multi-stage optimization. The result mixed integer programming (MIP) is NP-hard in general, thus we reformulate it as a sparse ℓ_0-minimization problem and then relax it to a linear programming (LP). Guided by the proposed optimization approach, we develop an efficient online resource allocation algorithm for practical systems, where a-priori knowledge of future service arrivals is not available. Finally, simulation results are provided to validate the proposed algorithms.

Keywords

Cellular/Infostation Integrated Network; On-Demand Multimedia Services; High-Speed Railway

Introduction

In recent years, the high-speed railway (HSR) has been developed rapidly as a fast, convenient and green public transportation system (J. Wang etc., 2014). There is a strong demand for various multimedia services such as online game and videophone, especially during a long time travelling (L. Tian etc., 2012; S. F. Xu etc., 2014). However, the data transmission rate degrades significantly when the train moves with an extremely high speed, e.g. 360 km/h, due to the serious Doppler shift, high penetration loss of signal into the carriage and varying channel conditions (Y. Zhao etc., 2012). Specifically, the passengers in the train will not enjoy the continuous multimedia (CM) services smoothly, since the CM services, such as IPTV and VoIP (O. B. Karimi etc., 2012), is transmitted as a continuous packet flow. Hence, how to deliver CM streaming effectively over limited resource and provide passengers with a high-quality end-user experience is an interesting and challenging problem in the HSR environment.

To address these challenges, we consider a cellular/infostation integrated HSR network architecture (H. Liang and W. Zhuang, 2012) in this work. In the integrated network, a cellular network with seamless coverage is supposed to support control channels for service requests and acknowledgements, while infostations deployed along the rail line can provide high-rate, high-quality and efficient data transmission. Based on this architecture, an optimal resource allocation problem for on-demand data delivery is investigated (H. Liang and W. Zhuang, 2012), which takes the intermittent network connectivity and multi-service demands into consideration. In (T. Chen etc., 2013) the total weight of received packets insteading of the fully completed requests was maximized. However, these works focused only in the data transmission integrity, with no attention to the continuity of data delivery, which will finally affect the playback continuity at passenger devices. For the best of our knowledge, the resource allocation problem that can jointly consider the integrity and continuity of on-demand data delivery in such an intermittently connected network has not been studied in the previous works.

In this paper, we investigate the downlink resource allocation problem in HSR downlink OFDMA (Orthogonal Frequency Division Multiple Access) system with a cellular/infostation integrated network architecture. The inter-carrier interference (ICI) due to the high moving speed is considered in characterizing the link capacity. Taking the train trajectory and CM

service demands into account, the resource allocation problem is formulated by a two-stage optimization programming, wherein the integrity of on-demand data delivery is considered in the first stage, while the continuity is considered in the second stage. Then we reformulate the two-stage optimization as a single-stage ℓ_0-minimization problem, and it is further approximated by a linear programming (LP). Moreover, we propose an efficient online resource allocation policy based on the insights gained from the offline analysis.

The rest of this paper is organized as follows. The system model and assumptions are described in Section 2, and the problem formulation is provided in Section 3. Section 4 presents the offline resource allocation algorithm and Section 5 proposes the online algorithm. Simulation results and discussions are given in Section 6, followed by the conclusions in Section 7.

Notation: $[x]^+ = \max\{x, 0\}$ and $\lfloor x \rfloor = \max\{n \in \mathbb{Z} | n \leq x\}$. $\mathbb{Z}^+$ represents the set of nonnegative integers. For given vector x, $\| x \|_0$ and $\| x \|_1$ denote the ℓ_0 and ℓ_1 norm, respectively.

System Model

As shown in Fig. 1, we consider a HSR downlink OFDMA system with a cellular/infostation integrated network architecture (H. Liang and W. Zhuang, 2012). H infostations with small coverage areas are deployed along the rail track to provide efficient data transmission, while the cellular network with seamless coverage is used for supporting the control channels over the region. A central controller (CC) is deployed to allocate the network radio resources for the whole network. The infostations can communicate with the vehicle station (VS) installed on the top of the train. Meanwhile, the VS is further connected to the access points (APs) inside the train. Assuming that the data transmission rate from the VS to passenger devices is sufficiently high, hence the data packet can be received successfully if it has been delivered to the VS. When the train is moving, the passengers send service requests from the VS to the content server (CS) through the cellular network. The requested data packets are then delivered from the CS to the VS via the infostations, and the VS will eventually forward the data to the passenger devices. For simplicity, we assume that the buffer space of VS is unlimited and the data service rate is equal to the end-user playback rate, which is assumed to be constant.

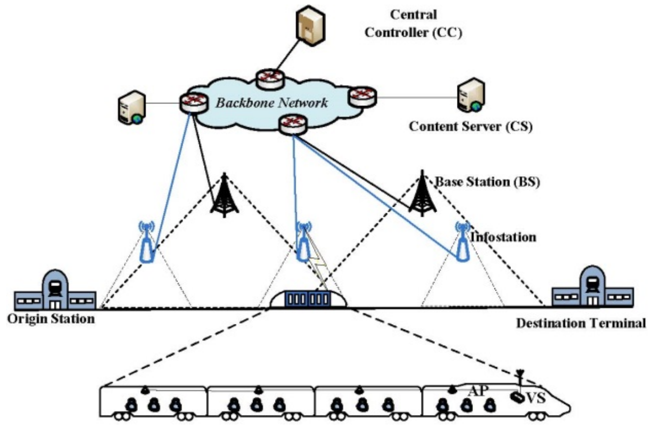

FIG. 1 SYSTEM MODEL

Note that each high-speed train moves on a predetermined rail line with highly stable time schedule $[T_I, T_O]$ from the origin station to its destination terminal. Thus the information of train trajectory and network resources can be obtained by the CC in advance with high accuracy (H. Liang and W. Zhuang, 2012; S. F. Xu etc., 2014). Each infostation with power P watts covers a segment of the rail line based on its wireless transmission range. Let T_h^i and T_h^o represent the time instants for the train to come into and go out of the coverage range of the hth infostation respectively, where $h \in \mathcal{H} \triangleq \{1,2, \ldots H\}$, and w.l.o.g., $T_1^i = T_I$ and $T_{H+1}^i = T_O$. Thus, the trip duration is partitioned into H non-overlapped periods according to the entry time, i.e., $[T_h^i, T_{h+1}^i]$ for all h. Then, each period is further partitioned into K_h slots with fixed length T_F, where $K_h = \lfloor (T_{h+1}^i - T_h^i)/T_F \rfloor$. Assume w.l.o.g. that $K_h = K, \forall h$. When the train visits the hth infostation, the kth ($k \in \mathcal{K}_h \triangleq \{1,2,\ldots,K\}$) slot begins and ends at times $T_h^i + (k-1)T_F$ and $T_h^i + kT_F$, respectively.

The data of CM services is encoded and segmented into a large number of packets, and we denote the link capacity A_{hk} as the maximum number of packets that can be delivered from the hth infostation to the VS at the kth slot, i.e.,

$$A_{hk} = \begin{cases} \lfloor R(t) T_F / L_p \rfloor, & \text{if } T_h^i \leq t < T_h^o, \\ 0, & \text{if } T_h^o \leq t < T_{h+1}^i, \end{cases} \quad (1)$$

where $t \triangleq T_h^i + (k-1)T_F, \forall k, h$, and we assume that each packet has equal size of L_p bits, and R(t) represents the maximum achievable data rate at slot t. For a downlink OFDMA system with N subcarriers, the R(t) is showed by

$$R(t) = \sum_{n=1}^{N} W \log_2 \left(1 + \frac{p_n(t) h_n^2(t)}{N_0 W + ICI_n(t)}\right), \quad (2)$$

where W denotes the bandwidth of the sub-carrier, N_0

represents the single-sided noise power spectral density (PSD) of each sub-carrier, $p_n(t)$ and $h_n(t)$ are the transmission power and channel gain of the nth subcarrier at the tth slot, respectively, and $\mathrm{ICI}_n(t)$ is the inter-carrier-interference experienced on the nth subcarrier at the tth slot.

The wireless channel is assumed to be frequency selective fading, where different sub-carriers will experience different channel gains, and the knowledge of channel state information is available at CC. Thus, the total power P can be allocated among the subcarriers by using waterfilling such that the rate can be maximized. Noticing that the ICI power $\mathrm{ICI}_n(t)$ caused by Doppler shift is not coordinated among different subcarriers, but its average impact is considered. By doing so, the wireless system can adapt to the varying channel timely and easily. According to (Y. Li and L. Cimini, 2001; T. Wang etc., 2006), the $\mathrm{ICI}_n(t)$ can be expressed as

$$\mathrm{ICI}_n(t) \approx \frac{(f_d T_s)^2}{2} \sum_{l=1, l\neq n}^{N} \frac{p_l(t)}{(l-n)^2}, \quad (3)$$

where T_s is the OFDMA symbol duration, and $f_d = vf_c/c$ represents maximum Doppler shift with v the moving speed, f_c the carrier frequency and c the velocity of light. A tight universal upper bound on the ICI power (Y. Li and L. Cimini, 2001) can be employed to ease the challenge in this paper

$$\mathrm{ICI}_{\mathrm{upperbound}} \leq \frac{1}{12}(2\pi f_d T_s \sqrt{P})^2. \quad (4)$$

Consider a single trip of a train from an origin station to a destination terminal. An on-demand CM service $s \in \mathcal{S} \triangleq \{1,2,\dots,S\}$ is requested at time G_s, and if $Q_s \in \mathbb{Z}^+$ packets are delivered to the VS beforetheir playback deadline D_s, a reward w_s will be obtained. Moreover, for a slotted transmission, the small difference between $G_s(D_s)$ and the beginning time of its generated slot is omitted. Since the data packets delivered before their playback deadline can be successfully decoded at passenger devices, the required minimum average service rate to play out a service s smoothly is showed by $V_s = Q_s/(D_s - G_s)$. In the sequel, we will thus investigate the optimal resource allocation for CM service scheduling.

Problem Formulation

In this section, we discuss an optimal resource allocation problem in the HSR downlink OFDMA system. The problem is formulated as a two-stage optimization programming.

Define x_{hks} as the packet number of service s delivered from the hth infostation at kth slot. Let ψ_s be a delivery indicator of service, which equals to 1 if service is completely delivered before its deadline, and 0 otherwise. Then ψ_s can be expressed by

$$\psi_s = \begin{cases} 1, & \text{if } \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \geq Q_s, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The first stage optimization problem with the goal of maximizing the total reward of delivered services over a trip of the train is then formulated as

$$\begin{aligned} (\mathbf{P1}) \quad & \max_{\{x_{hks}\}} \sum_{s\in\mathcal{S}} w_s \psi_s & (6a) \\ \text{s.t.} \quad & x_{hks} \in \mathbb{Z}^+, \ \ \forall k \in \mathcal{K}_h, h \in \mathcal{H}, s \in \mathcal{S} & (6b) \\ & \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \leq Q_s, \ \ \forall s \in \mathcal{S} & (6c) \\ & \sum_{s\in\mathcal{S}} x_{hks} \leq A_{hk}, \ \ \forall k \in \mathcal{K}_h, \ h \in \mathcal{H} & (6d) \\ & x_{hks} = 0, \ \ \text{if } G_s \geq T_h^i + kT_F \\ & \qquad \text{or } D_s \leq T_h^i + (k-1)T_F, \\ & \qquad \forall k \in \mathcal{K}_h, \ h \in \mathcal{H}, \ \ s \in \mathcal{S} & (6e) \end{aligned} \quad (6)$$

where the constraint (6b) means that negative resource allocation is not allowed, (6d) states that the packet number that can be delivered from hth infostation at kth slot is limited by the capacity A_{hk}, while (6e) shows that the packets can only be delivered between the request time and the deadline.

The problem P1 will determine the optimal service set (with maximum total reward), denoted by $\mathcal{S}^*$, that can be completely delivered before their deadlines. Notice that $\mathcal{S}^*$ might not be unique. And moreover, P1 can only guarantee the data transmission integrity, without considering of the continuity of data streams as well as the arrival order. As we known that because data delivery over the wireless links with limited resource and intermittent connectivity, the strict continuity constraint is prohibitively difficult to meet. But on the other side, due to the human perceptual inability to notice small deviations of presentation rate, people are more sensitive to a low-frequency but long-lasting disruptions than to a high-frequency but short-lived disruptions (M. C. Yuang etc., 1996). As a result, a better perceptual quality can be expected by replacing large continuity disruptions with shorter ones (N. Laoutaris and I. Stavrakakis, 2002), and it is beneficial to minimize the amount of cumulative discontinuous packets during each slot to improve the perceptual quality. The cumulative discontinuous means that the amount of available data packets

within each slot is less than that consumed by playback at the same time. Accordingly, the cumulative discontinuous packet number of service at the k th slot when the train is served by the hth infostation can be formulated by

$$f(x_{hks}) \triangleq V_s \Delta t_{hks} - \left(\sum_{m=1}^{h-1} \sum_{t=1}^{K_{h-1}} x_{mts} + \sum_{t=1}^{k} x_{hts} \right), \tag{7}$$

where$\Delta t_{hks} = \left[\min\{T_h^i + (k-1)T_F, D_s\} - G_s\right]^+$ denotes the transmission duration of the service from G_s to the kth slot within the hth infostation, $V_s \Delta t_{hks}$ and $\sum_{m=1}^{h-1}\sum_{t=1}^{K_{h-1}} x_{mts} + \sum_{t=1}^{k} x_{hts}$ refer to the required packet number for continuous playback of multimedia data and actual transmitted packet number during Δt_{hks} for service, respectively.

Then, the second stage problem with the objective of minimizing the weighted total number of cumulative discontinuity packets of the delivered services in $\mathcal{S}^*$ over a trip of the train is formulated by

$$\begin{aligned} (\mathbf{P2}) \min_{\{x_{hks}\}_{s\in\mathcal{S}^*}} \quad & \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} \sum_{s\in\mathcal{S}^*} w_s f(x_{hks}) && (8a) \\ \text{s.t.} \quad & x_{hks} \in \mathbb{Z}^+, \quad \forall k \in \mathcal{K}_h, h \in \mathcal{H}, s \in \mathcal{S}^*, && (8b) \\ & \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \le Q_s, \quad \forall s \in \mathcal{S}^*, && (8c) \\ & \sum_{s\in\mathcal{S}^*} x_{hks} \le A_{hk}, \quad \forall k \in \mathcal{K}_h, \forall h \in \mathcal{H}, && (8d) \\ & x_{hks} = 0, \quad \text{if } G_s \ge T_h^i + kT_F \\ & \qquad \text{or } D_s \le T_h^i + (k-1)T_F, \\ & \qquad \forall k \in \mathcal{K}_h, \forall h \in \mathcal{H}, \forall s \in \mathcal{S}^*. && (8e) \end{aligned} \tag{8}$$

The potential rational under the above formulation is to facilitate data transmitting, and if a data packet is late, it should be transmitted as early as possible, instead of being discarded, to protect the continuity of the stream from further degradation. Moreover, services with higher reward values receive higher priority to obtain enough packets for maintaining continuous playback.

Offline Resource Allocation Algorithm

As we known, two-stage optimization is, in general, significantly harder than single-stage optimization. In this section, we will reformulate the two-stage problem P1-P2 into an equivalent single-stage optimization problem by fully exploiting the inherent structure.

2-in-1 Reformulation

To proceed, let us define the weighted reward matrix $\mathbf{W}$ as $\mathbf{e}[\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_S}]$, where $\mathbf{e} \in \mathbb{R}^{HK_h}$ represents the all-one vector. The packet allocation matrix is denoted by $\mathbf{X} \in \mathbb{R}^{HK_h \times S}$ with its entry$[\mathbf{X}]_{hk_h,s} = x_{hk_hs}$. Then we have

$$\sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} \sum_{s\in\mathcal{S}} w_s \left(\sum_{m=1}^{h-1} \sum_{t=1}^{K_{h-1}} x_{mts} + \sum_{t=1}^{k} x_{hts} \right) = \operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}),$$

Where $\mathbf{B}$ is a well-defined $(HK_h \times HK_h)$-dimensional unit lower triangular matrix. Meanwhile, we note that $\sum_{h\in\mathcal{H}}\sum_{k\in\mathcal{K}_h}\sum_{s\in\mathcal{S}} w_s V_s \Delta t_{hks}$ is a constant. Thus the problem P2 is equivalent to

$$\begin{aligned} \max_{\{x_{hks}\}_{s\in\mathcal{S}^*}} \quad & \operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}) \\ \text{s.t.} \quad & (8b) - (8e). \end{aligned} \tag{9}$$

Introduced a parameter α for the two-stage programming, then we have

$$\begin{aligned} (\mathbf{P3}) \quad \max_{\{x_{hks}\}} \quad & \sum_{s\in\mathcal{S}} w_s \psi_s + \alpha \operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}) \\ \text{s.t.} \quad & (6b) - (6e), \end{aligned} \tag{10}$$

Where α is a constant satisfying $0 < \alpha < \frac{w^\star}{\operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}^*)}$, $w^\star = \min\{w_1, w_2, \cdots w_S\}$, and $\mathbf{X}^*$ is the solution to problem P3. Actually, if there are multiple solutions to P1, the above formulation is capable of picking the one with minimum weighted total cumulative discontinuity packets. We can draw the following equality between P1-P2 and P3.

Lemma 1. The two-stage optimization problem P1-P2 is equivalent to the problem P3.

Proof. The proof of Lemma 1 follows (Y. F. Liu etc., 2012) and is omitted for brevity.

The problem P3 is a mixed integer nonlinear programming (MINLP) problem, which in general NP-hard (L. A. Wolsey, 1998). Now, we can transform the single-stage problem P3 to an ℓ_0 -norm minimization problem. Let $\phi_s = 1 - \psi_s$ and we have $\phi_s = \left\|Q_s - \sum_{h\in\mathcal{H}}\sum_{k\in\mathcal{K}_h} x_{hks}\right\|_0$.

Thus the problem P3 is equivalent to the following sparse optimization problem

$$\begin{aligned} (\mathbf{P4}) \min_{\{x_{hks}\}} \quad & \sum_{s\in\mathcal{S}} w_s \left\| Q_s - \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \right\|_0 - \alpha \operatorname{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}) \\ \text{s.t.} \quad & (6b) - (6c). \end{aligned} \tag{11}$$

Now, it can be clearly seen that the difficulty of (P4) comes from the optimization over integer variables and the non-convex objective function involving ℓ_0-minimization, which is NP-hard in general (L. A. Wolsey, 1998). Therefore, a suboptimal algorithm is

proposed in the following subsection to solve this problem efficiently.

Linear Programming Approximate (LPA)

First of all, since ℓ_0-optimization problem P4 is NP-hard, it is natural to consider its ℓ_1-convex relaxation. By relaxing the hard constraints $x_{hks} \in \mathbb{Z}^+$ to linear inequality constraints $0 \le x_{hks} \le Q_s$, the final approximation of the problem P4 can be expressed as

$$\begin{aligned} \min_{\{x_{hks}\}} & \sum_{s\in\mathcal{S}} w_s \left\| Q_s - \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \right\|_1 - \alpha \mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}) \\ \text{s.t.} \quad & 0 \le x_{hks} \le Q_s, \forall k \in \mathcal{K}_h, \forall h \in \mathcal{H}, \forall s \in \mathcal{S}, \\ & (6c) - (6e), \end{aligned} \tag{12}$$

However, it can be easily verified that the case $\sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} > Q_s$ doesn't happen, because the resources are underutilized by delivering more than Q_s packets for service s. Therefore, (12) is equivalent to the following LP problem

$$\begin{aligned} \min_{\{x_{hks}\}} & \sum_{s\in\mathcal{S}} w_s \left(Q_s - \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} x_{hks} \right) - \alpha \mathrm{tr}(\mathbf{W}^{\mathrm{T}}\mathbf{B}\mathbf{X}) \\ \text{s.t.} \quad & 0 \le x_{hks} \le Q_s, \forall k \in \mathcal{K}_h, \forall h \in \mathcal{H}, \forall s \in \mathcal{S}, \\ & (6c) - (6e), \end{aligned} \tag{13}$$

which is a convex optimization problem and the optimum solution x^*_{hks} can be obtained by CVX. Since x^*_{hks} is not necessarily a nonnegative integer, we need more steps to attain a feasible solution. First, the solution x^*_{hks} is rounded down to the nearest integer $\tilde{x}^*_{hks} = \lfloor x^*_{hks} \rfloor$. Then, based on the insight discussed in section 3, a packet loading scheme is detailed in Algorithm 1.

In step 1, the remaining packet number of service s and the remaining capacity of the kth slot from the hth infostation is showed respectively by

$$\tilde{Q}_s = Q_s - \sum_{h\in\mathcal{H}} \sum_{k\in\mathcal{K}_h} \tilde{x}^*_{hks}, \tag{14}$$

$$\tilde{A}_{hk} = A_{hk} - \sum_{s\in\mathcal{S}} \tilde{x}^*_{hks}. \tag{0.0.15}$$

In step 5, $\mathcal{S}^c_{hk}$ represents the active service set which can possibly be allocated with more packets at the kth slot of the hth infostation, showed by

$$\mathcal{S}^c_{hk} = \left\{ s \,\middle|\, s \in S, \tilde{Q}_s > 0, G_s < T^i_h + kT_F, D_s > T^i_h + (k-1)T_F \right\} \tag{16}$$

Algorithm 1 iteratively allocates the remaining capacity $\tilde{A}_{hk}$ to the active services in descending order of $w_s/\tilde{Q}_s$, until the capacity of the slot is fully utilized. Based on these principles, the complete description of the proposed linear programming approximate (LPA) algorithm is showed in Algorithm 2.

Algorithm 1 Packet Loading Scheme

1: **Initialize** $\tilde{A}_{hk}, \tilde{Q}_s, \tilde{x}^*_{hks}, \forall h \in \mathcal{H}, k \in \mathcal{K}_h, s \in \mathcal{S}$;
2: **for** $h = 1$ to H **do**
3:   **for** $k = 1$ to K_h **do**
4:     **while** $\tilde{A}_{hk} > 0$ **do**
5:       Calculate $\mathcal{S}^c_{hk}$ by (16);
6:       **if** $\mathcal{S}^c_{hk} = \emptyset$ **then**
7:         break;
8:       **else**
9:         $s^* = \arg\max_{s\in S_A} \{w_s/\tilde{Q}_s\}$;
10:         $\tilde{x}^*_{hks*} := \tilde{x}^*_{hks*} + 1$;
11:         $\tilde{Q}_{s*} := \tilde{Q}_{s*} - 1$;
12:         $\tilde{A}_{hk} := \tilde{A}_{hk} - 1$;
13:       **end if**
14:     **end while**
15:   **end for**
16: **end for**

Algorithm 2 Proposed LP Approximate Algorithm

1: Initialize $P, N, W, N_0, T_s, f_d, h_n(t), Q_s, G_s, D_s$ and w_s, $\forall t \in [T_I, T_O], \forall h \in \mathcal{H}, k \in \mathcal{K}_h, s \in \mathcal{S}$;
2: Calculate $\{A_{hk}\}$ by water-filling;
3: Resource allocation by solving (13).
4: Let $\tilde{x}^*_{hks} = \lfloor x^*_{hks} \rfloor$ and call Alg. 1 for packet loading.

Online Resource Allocation Algorithm

In this section, we will study the online version of the offline algorithm that presented in the section 4. Supposing that the knowledge of the future service arrivals is not available, and only channel information per slot within a channel coherence interval is known at infostations. The main idea behind our online algorithm is to allocate the packets according to the offline policy based on the current service arrivals and channel capacities within the channel coherence interval, and then reallocate when new services arrive.

Let L represent the number of slots in a channel coherence interval. Assume that a set S_0 packets arrive at time instant 1, and meanwhile L slots channel capacities are available. Then we construct a (sub-) system over the slots $[1, L]$. For such a subsystem, we run the proposed offline algorithm to find the optimal packet schedule policy. The packet delivery follows this policy until new services arrive at t. Then we treat the current time instant t as new "1" instant, and update the set S_0 including the newly arriving services, and removing the old services whose deadlines are past or packets all delivered. For other old services, we update their D_s and Q_s as $D_s - t$ and the number of their remaining packets, respectively. Provided by the new L slots channel information, the offline algorithm is runnig for the new subsystem

$[t, t + L - 1]$, and the subsequent packet delivery follows the resultant new policy. This process continues until all the packets are delivered.

Simulation Results

In this section, we evaluate the performance of the proposed resource allocation algorithms. The cellular/infostation integrated network under simulation assumes that there are $H = 5$ infostations, each with the transmission range 100 meters. An HSR train moves at the speed of 360 km/h and all wireless links experience Rician fading, where the channel gain $h_n(t)$ is generated by Rician distribution with K-factor 1. The duration of a time slot is 20 ms. The data packet size is $L_p = 500$ bits. Moreover, the size of each service (Q_s) is uniformly distributed within [25, 50] kbits. The service requests follow the Poisson process with average service arrival rate λ, and the lifetime ($D_s - G_s$) of each service is exponentially distributed with average value 10 seconds. The reward of each service (w_s) is uniformly distributed within [1,10].The other simulation parameters are set as below: $T_s = 1.33 \times 10^{-4}$ s, $f_c = 3.5$ GHz, $W = 7.5$ KHz, $N_0 = 2s10^{-7}$ W/Hz, $P = 20$ dBm, $L = 4$, $N = 4$, $\epsilon = 0.1$, and the maximum number of iterations $n_{\max} = 10$.

We compare the performance of the proposed offline and online algorithms with three existing algorithms: first-in-first-out (FIFO), earliest due date (EDD) and the Smith ratio based algorithm in (H. Liang and W. Zhuang, 2012). All figures are obtained by averaging over 500 independent simulation runs.

Fig. 2 illustrates the relationship between the total reward of delivered services over the trip of the train and the average service arrival rate (λ).It can be seen that, by incorporating the demands of multiple CM services and train trajectory, our proposed offline resource allocation algorithm achieves the highest total reward.It is also interesting to note that the proposed online algorithm performs close to the offline algorithm and outperforms the other three algorithms.Moreover, as shown in Fig. 3, the weighted total number of cumulative discontinuity packets increases as λ grows from 0.5 to 1.2 . For a larger number services, less resources can be allocated to deliver each service. As a result, a less number of services can be delivered smoothly under the limited resources. It can be observed that the cumulative discontinuity packets of the existing algorithms, which don't take the continuity of data streams into consideration, are much more than those of the proposed algorithms. That means the proposed algorithms achieve better perceptual quality than the existed algorithms.

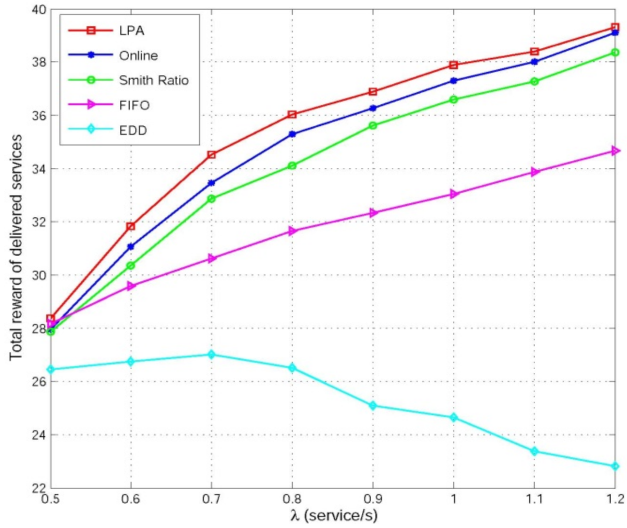

FIG. 2 TOTAL REWARD OF DELIVERED SERVICES VERSUS λ

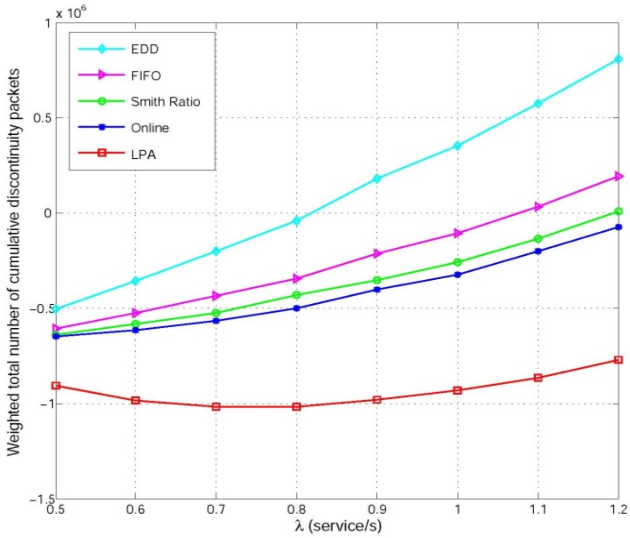

FIG. 3 WEIGHTED TOTAL NUMBER OF CUMULATIVE DISCONTINUITY PACKETS VERSUS λ

From Fig. 2 and 3, we can see that the proposed resource allocation algorithms can obtain more rewards but less cumulative discontinuity packets than the existing algorithms under different system parameters.

Conclusions

In this paper, we investigate the optimal resource allocation for on-demand continuous multimedia service delivery in a cellular/infostation integrated network for high-speed railways. A two-stage optimization problem is formulated by taking account of the integrity and continuity of the on-demand data delivery. To address the NP-hard mixed integer

programming, an LP approximation algorithm has been developed. Guided by the optimal offline strategy, we proposed an online algorithm for practical systems where a-priori knowledge of the future service arrivals is not available. Finally, the simulation results demonstrate the effectiveness of the proposed algorithms.

ACKNOWLEDGMENT

This work is supported by the Fundamental Research Funds for the Central Universities (2014JBM153), the China Postdoctoral Science Foundation (No. 2013M530519), the State Key Lab of Rail Traffic Control and Safety (No. RCS2012ZZ004, RCS2014ZQ004), the Natural Science Foundation of China (No. U1334202) and the Key Grant Project of Chinese Ministry of Education (No. 313006). The authors would like to thank Prof. Xia Chen for her valuable suggestions. Chao Shen is the corresponding author.

REFERENCES

H. Liang and W. Zhuang, "Efficient on-demand data service delivery to high-speed trains in cellular/infostation integrated networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 780–791, May 2012.

J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 675–683, 2012.

L. A. Wolsey, *Integer programming*.New York: Wiley-Interscience, 1998.

L. Tian, J. Li, Y. Huang, J. Shi, and J. Zhou, "Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 708–718, 2012.

M. C. Yuang, S. T. Liang, Y. G. Chen, and C. L. Shen, "Dynamic video playout smoothing method for multimedia applications," in *Proc. IEEE ICC*, vol. 3, 1996, pp. 1365–1369.

N. Laoutaris and I. Stavrakakis, "Intrastream synchronization for continuous media streams: A survey of playout schedulers," *IEEE Netw.*, vol. 16, no. 3, pp. 30–40, 2002.

O. B. Karimi, J. Liu, and C. Wang, "Seamless wireless connectivity for multimedia services in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 729–739, 2012.

S. F. Xu, G. Zhu, C. Shen, Y. Lei, and Z. D. Zhong, "Delay-aware online service scheduling in high-speed railway communication systems," *Math. Probl.Eng.*, vol. 2014, Article ID 248174, 10 pages, 2014.

S. F. Xu, G. Zhu, C. Shen, Y. Lei, and Z. Zhong, "Utility-based resource allocation in high-speed railway wireless networks," *EURASIP J. WIREL. COMM.*, vol. 2014, no. 1, pp. 1–14, 2014.

T. Chen, H. Shan, and X. Wang, "Packet scheduling for on-demand data services to high-speed trains over wireless links," in *Proc. IEEE Globecom*, 2013.

T. Wang, J. Proakis, E. Masry, and J. Zeidler, "Performance degradation of OFDM systems due to doppler spreading," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1422–1432, 2006.

Y. F. Liu, Y. H. Dai, and Z. Q. Luo, "Joint power and admission control via linear programming deflation," in *Proc. ICASSP*, 2012, pp. 2873–2876.

Y. Li and L. Cimini, "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 401–404, 2001.

Y. Zhao, X. Li, X. Zhang, Y. Li, and H. Ji, "Multidimensional resource allocation strategy for high-speed railway MIMO-OFDM system," in *Proc. IEEE Globecom*, 2012, pp. 1653–1657.